GAUSS-NEWTON OPTIMIZATION IN DIFFEOMORPHIC REGISTRATION

Monica Hernandez and Salvador Olmos

Communication Technologies Group (GTC) and
Aragon Institute of Engineering Research (I3A)
University of Zaragoza, Spain

ABSTRACT

In this article, we propose a numerical implementation of Gauss-Newton's method for optimization in diffeomorphic registration in the Large Deformation Diffeomorphic Metric Mapping framework. The computations of the Gâteaux derivatives of the objective function are performed in the tangent space of the Riemannian manifold of diffeomorphisms. The resulting algorithm has been compared to gradient descent optimization in brain MRI anatomical images. The experiments have shown similar accuracy for both techniques at steady-state while Gauss-Newton has resulted to be more robust with a faster rate of convergence.

Index Terms— Diffeomorphic registration, optimization methods, Gauss-Newton, Hilbert spaces

1. INTRODUCTION

Image registration is usually defined as a variational problem involving the metric that measures the image matching after registration and a regularization constraint in the geometric transformation that maps the source into the target. Numerical optimization techniques play a crucial role in order to find the optimal transformation that minimizes the objective function [1].

In diffeomorphic registration transformations are usually assumed to belong to a Riemannian manifold of diffeomorphisms. The majority of the approaches have focused on the characterization of the diffeomorphic transformations in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [2, 3, 4]. Much less attention has been paid to the optimization strategy where classical gradient descent method is often used.

Recently, Ashburner et al. have proposed a numerical implementation of Gauss-Newton's method for the LDDMM variational problem [5]. The computations of the Gâteaux derivatives of the objective function are performed in the space of L^2-functions. In consequence, the action of the linear operator involved in the regularization term has to be formulated using the matrix representation of the convolution. As a result, the algorithm results into a high dimensional matrix inversion problem [1]. Although there exist well known multi-grid techniques to numerically solve these problems [6], the memory requirements for diffeomorphic registration hinders their execution in standard machines. Moreover, multi-grid schemes need the definition of the injection and the interpolation operators associated to the elements involved in the registration in order to compute fine-to-coarse and coarse-to-fine samplings. In the case of images, downsampling and linear interpolation are good candidates for these operators. In the case of diffeomorphic transformations, however, the definition of these operators remains an open question.

In this article we propose an alternative numerical implementation of Gauss-Newton's method for LDDMM registration. The computations of the Gâteaux derivatives of the objective function are performed in the tangent space of the Riemannian manifold of diffeomorphisms. This way, the action of the linear operator involved in the regularization term can be directly formulated using convolution. As a result, this algorithm results into a 3 dimensional problem for each point in the computation domain and can be executed in standard machines. The resulting algorithm has been compared to gradient descent optimization in a database of brain MRI anatomical images. The experiments have shown similar accuracy for both techniques at steady-state while Gauss-Newton has resulted to be more robust with a faster rate of convergence.

The rest of the article is divided as follows. In Section 2 we briefly revisit the framework for diffeomorphic registration and derive the computations for Gauss-Newton's method. In Section 3 we present and discuss the experimental results. Finally, Section 4 presents some concluding remarks.

2. METHOD

2.1. Diffeomorphic registration

In LDDMM diffeomorphic registration, transformations are usually assumed to belong to a group of diffeomorphisms $Diff(\Omega)$ (i.e. differentiable maps $\varphi : \Omega \rightarrow \Omega$ with differentiable inverse) endowed with a Hilbert differentiable struc-

This work has been partially supported by Spanish research grants TEC2006-13966-C03-02 and FIS PI04/1795.

ture of Riemannian manifold. The tangent space V is the set of Sobolev class vector fields in Ω. The Riemannian metric is defined from the scalar product $\langle v, w\rangle_V = \langle Lv, Lw\rangle_{L^2}$ where L is a linear invertible differentiable operator.

Diffeomorphic registration from a template image I_0 to a target I_1 is represented by the end point $\varphi = \phi(1)$ of a path of diffeomorphisms $\phi(t)$ starting at the identity element and resulting from the minimization of the energy functional

$$E(\varphi) = E(\phi)^2 + \frac{1}{\sigma^2}\|I_0 \circ \varphi^{-1} - I_1\|^2_{L^2} \tag{1}$$

where the term $E(\phi)^2$ imposes a regularization in the path energy and the L^2 norm measures the matching between the images after registration. The factor $1/\sigma^2$ balances the energy contribution between regularization and matching.

In this article, diffeomorphic transformations are parameterized using stationary vector fields [7, 4]. The stationary parameterization is closely related to the group structure defined in $Diff(\Omega)$ as the paths starting at the identity that can be parameterized using stationary vector fields are exactly the one-parameter subgroups. Diffeomorphisms belonging to one-parameter subgroups can be computed from the group exponential map $\text{Exp} : V \rightarrow Diff(\Omega)$. The path energy is defined from the metric associated to the infinitesimal generator $E(\phi)^2 = \|w\|^2_V$, where $\phi(1) = \text{Exp}(w)$. Therefore, diffeomorphic registration is posed from the minimization of the energy functional

$$E(w) = \|w\|^2_V + \frac{1}{\sigma^2}\|I_0 \circ \text{Exp}(w)^{-1} - I_1\|^2_{L^2} \tag{2}$$

2.2. Gauss-Newton's method

Let $E(w)$ be a twice Frechet differentiable energy functional defined in a general convex vector space V. Second order optimization techniques are obtained from the second order Taylor approximation

$$E(w+h) \cong E(w) + \langle \nabla_w E(w), h\rangle_V + \frac{1}{2}\langle h, H_w E(w) h\rangle_V \tag{3}$$

where $\nabla_w E(w)$ and $H_w E(w)$ are the Frechet differentials defined in V.

Newton's method is the most popular second order optimization technique. The minimization of $E(w)$ is approached with an iterative scheme

$$w^{k+1} = w^k - \epsilon \cdot H_w E(w^k)^{-1} \cdot \nabla_w E(w^k) \tag{4}$$

where the search direction exploits not only the gradient but also the Hessian structure thus providing a faster rate of convergence than methods just based on the gradient.

The Gâteaux derivative of an energy functional $E(w)$ along $h \in V$ is defined as its variation under the perturbation of w in the direction of h, $\partial_h E(w)$. The second order Gâteaux derivative, $\partial_{hh} E(w)$ is computed recursively from the first order derivative. In Frechet spaces, the gradient operator relates the Frechet differential and the Gâteaux derivative (whenever both derivatives exist) by

$$\partial_h E(w) = \langle \nabla_w E(w), h\rangle_V \tag{5}$$

The Hessian operator relates the second order Frechet differential and the second order Gâteaux derivative by

$$\partial_{hh} E(w) = \langle h, H_w E(w)\ h\rangle_V \tag{6}$$

Using Gâteaux derivatives, the computations of Newton's equations for the energy functional defined in Equation 2 can be derived. For simplicity, we divide $E(w) = E_1(w) + \frac{1}{\sigma^2}E_2(w)$ where $E_1(w) = \|w\|^2_V$ and $E_2(w) = \|I_0 \circ \text{Exp}(w)^{-1} - I_1\|^2_{L^2}$, and compute $\partial_h E_i(w)$ and $\partial_{hh} E_i(w)$, $i = 1, 2$ separately. Straightforward computations provide the derivatives related to E_1

$$\partial_h E_1(w) = 2\langle w, h\rangle_V \quad \partial_{hh} E_1(w) = 2\langle h, h\rangle_V \tag{7}$$

The first and second order variations of $E_2(w)$ in the direction of h are computed using the chain rule in L^2 and projecting the results into the space V using the inverse of operator L. In the computations, the Gâteaux derivative of the exponential map is computed using a first order approximation [4]. In order to simplify the notation, let denote with $r = I_0 \circ \text{Exp}(w)^{-1} - I_1$ the residual function associated to $E_2(w)$, $J = \nabla(I_0 \circ \text{Exp}(w)^{-1})$ the gradient and $H = \text{Hess}(I_0 \circ \text{Exp}(w)^{-1})$ the Hessian matrix associated to the transformed template. With this notation,

$$\begin{aligned} \partial_h E_2(w) &= -2\langle r\ J, h\rangle_{L^2} = -2\langle (L^\dagger L)^{-1}(r\ J), h\rangle_V \\ \partial_{hh} E_2(w) &= 2\langle h, (J^T \cdot J + r\ H)\cdot h\rangle_{L^2} \\ &= 2\langle h, (L^\dagger L)^{-2}(J^T \cdot J + r\ H)\cdot h\rangle_V \end{aligned} \tag{8}$$

and Newton's equations can be obtained from the combination of Equation 4 to Equation 8.

In practice, the computation of the Hessian $H_w E(w) = 2I_{\mathbb{R}^3} + \frac{2}{\sigma^2}(L^\dagger L)^{-2}(J^T \cdot J + r\ H)$ often leads to numerical problems as the term $J^T \cdot J + r\ H$ is not always guaranteed to be positive definite during optimization. Gauss-Newton's method is often used as a simplification of Newton's method that overcomes this limitation using a linear approximation of this term. Thus, Gauss-Newton's method for the minimization of $E(w)$ is approached with the iterative scheme

$$\begin{aligned} w^{k+1} = w^k - \epsilon \quad \cdot \quad & \left(2\ I_{\mathbb{R}^3} + \frac{2}{\sigma^2}(L^\dagger L)^{-2}(J^T \cdot J)\right)^{-1} \cdot \\ & \left(2\ w - \frac{2}{\sigma^2}(L^\dagger L)^{-1}(r\ J)\right) \end{aligned} \tag{9}$$

3. RESULTS AND DISCUSSION

In this section we have evaluated the influence of optimization in LDDMM diffeomorphic registration (Equation 2). To this

end, we have compared Gauss-Newton's method (GN) with gradient descent (GD) [4] in terms of accuracy, robustness, and efficiency. The study has been performed in a population of 18 T1-MRI brain anatomical images. Optimization results have been analyzed at three different resolution levels. In order to make results comparable, the same initialization ($w^{(0)} = 0$) has been selected for each level.

The iterative schemes corresponding to GN and GD can be included into the general formulation $w^{k+1} = w^k - \epsilon \cdot d(w^k)$ where $d(w^k)$ is the search direction at iteration k and ϵ controls the step size made along this direction. During optimization, the selection of the parameter ϵ is critical as it can determine the convergence of the algorithm. For example, if ϵ is selected to be too big, the step made along the search direction may not provide a step minimizing the energy. Otherwise, if ϵ is selected to be too small the algorithm may get trapped into a local minimum far from the global minimum.

In these experiments, a backtracking line-search strategy starting from an initial guess ϵ_0 has been used. In each iteration the step size ϵ has been selected to be the first parameter that provides a sufficient decrease in the energy according to Armijo's condition [8]. In the case of GN method, ϵ_0 has been selected to be equal to 1. Thus, GN optimization shows a superlinear rate of convergence. In GD method, the search of ϵ can be quite expensive, specially if the guess for ϵ_0 is far from the optimal value. For this reason, the value of ϵ_0 for the current iteration has been selected to be the estimated ϵ in the previous iteration.

In practice, the algorithms are considered to converge according to criteria that compromises between the number of iterations and maximum energy decrease. In these experiments, however, the algorithms are considered to converge only if Armijo's condition does not hold for any big enough ϵ or the number of elapsed iterations exceeds a threshold (in our case, 1000). Thus, convergence properties of both optimization methods can be fully studied.

3.1. Accuracy

Table 1 shows the number of iterations, the values of the energies involved in the objective function (Equation 2) and the relative sum of squared differences between the images at convergence. Table 2 shows the extrema of the Jacobian determinant of the resulting diffeomorphic transformation. Figure 1 shows the average curve of the image matching during optimization. The bars indicate the standard deviation through the database of patients. From these results, it can be seen that both optimization techniques provide similar registration results at convergence although GN shows a faster convergence rate than GD.

3.2. Robustness

Figure 2 shows some representative examples of the image matching curve during optimization. While GN shows a monotone decreasing convergence, GD is frequently prone to get trapped into local minima during a considerable number of iterations. Besides, GN shows a faster convergence rate than GD specially at the coarsest resolution levels and the initial iterations of the algorithm. Therefore, GN results a more appropriate optimization technique for diffeomorphic registration than GD, specially for multiresolution versions of the algorithm.

Table 1. Average and standard deviation of the number of iterations, # it, the energy associated to the diffeomorphic transformation, $\|\cdot\|_V$, the image matching, $\|\cdot\|_{L^2}$, and the relative sum of squared differences, RSSD, at convergence. Level 0 corresponds to the finest resolution level.

	level	# it	$\|\cdot\|_V$	$\|\cdot\|_{L^2}$	RSSD (%)
GN	2	170.58 ± 10.42	63.25 ± 5.30	10.49 ± 1.65	16.35 ± 1.01
	1	793.52 ± 37.15	258.48 ± 12.77	8.74 ± 1.06	11.37 ± 0.28
	0	777.29 ± 26.25	516.34 ± 13.68	14.17 ± 0.77	14.06 ± 1.32
GD	2	222.11 ± 10.79	63.52 ± 6.08	10.78 ± 1.79	16.29 ± 1.22
	1	896.58 ± 58.45	259.23 ± 14.20	8.69 ± 1.02	11.31 ± 0.33
	0	648.00 ± 31.62	499.68 ± 15.38	12.42 ± 0.91	13.36 ± 1.23

Table 2. Average and standard deviation of the extrema and the 0.2% quantiles of the Jacobian determinant associated to diffeomorphism φ at convergence. Level 0 corresponds to the finest resolution level.

	level	J_{min}	$J_{0.2\%}$	J_{max}	$J_{99.8\%}$
GN	2	0.04 ± 0.02	0.14 ± 0.07	10.78 ± 4.24	3.55 ± 0.61
	1	0.03 ± 0.01	0.12 ± 0.02	15.39 ± 12.07	3.62 ± 0.67
	0	0.08 ± 0.03	0.22 ± 0.06	11.05 ± 12.59	2.8 ± 0.38
GD	2	0.05 ± 0.03	0.14 ± 0.08	10.82 ± 4.13	3.61 ± 0.62
	1	0.04 ± 0.01	0.12 ± 0.03	14.87 ± 11.93	3.52 ± 0.63
	0	0.11 ± 0.02	0.27 ± 0.03	8.25 ± 8.80	2.74 ± 0.28

3.3. Efficiency

Memory requirements for GN and GD for a volume of size $155 \times 205 \times 170$ are equal to 1027 and 802 MB, respectively. Figure 3 shows the average curve of the elapsed time for each iteration. The bars indicate the standard deviation through the database of patients. Although the computation of the search direction $d(w^k)$ is more expensive for GN than for GD (31.92 vs 20.44 seconds in the finest resolution level in a 2.336 GHz machine), the line search results more expensive for GD, specially at the initial iterations of the algorithm. On the whole, GN results into a much more efficient implementation.

4. CONCLUSION

We have presented a numerical implementation of Gauss-Newton's method for diffeomorphic registration. The algorithm has been compared to gradient descent in a database of T1-MRI brain anatomical images. The results have shown that both algorithms provide similar accuracy while Gauss-Newton is more robust and efficient. Hence Gauss-Newton provides a more appropriate optimization technique for diffeomorphic registration, specially for multiresolution versions of the algorithm.

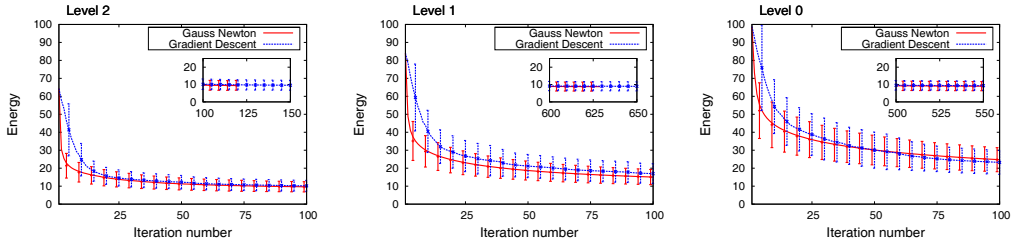

Fig. 1. Average and standard deviation of the image matching curve, $\|I_0 \circ \varphi^{-1} - I_1\|^2_{L^2}$, during optimization.

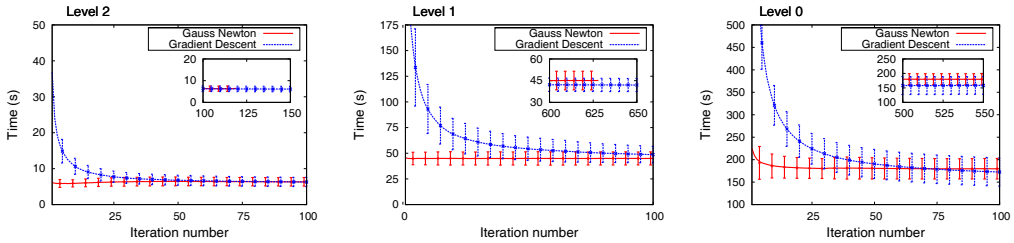

Fig. 3. Average and standard deviation of the time curve during optimization.

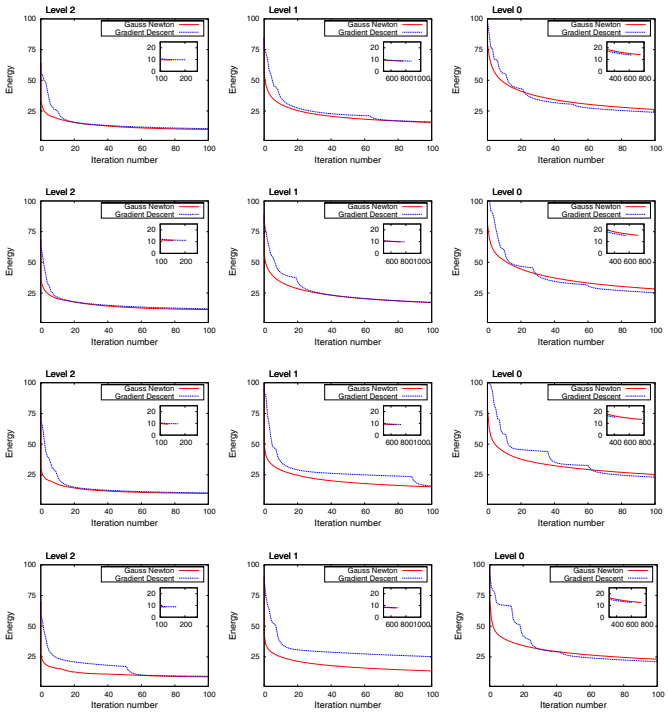

Fig. 2. Representative examples of the image matching curves during optimization.

5. REFERENCES

[1] J. Modersitzki, "Numerical methods for image registration," *Oxford University Press, New York*, 2004.

[2] M. F. Beg, M. I. Miller, A. Trouve, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vis.*, vol. 61 (2), pp. 139–157, 2005.

[3] L. Younes, "Jacobi fields in groups of diffeomorphisms and applications," *Quart. Appl. Math.*, vol. 65, pp. 113 – 134, 2007.

[4] M. Hernandez, M. N. Bossa, and S. Olmos, "Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields," *IEEE workshop on Math. Meth. in Biom. Image Anal. (MMBIA 07)*, 2007.

[5] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38(1), pp. 95 – 113, 2007.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipes in C: the art of scientific computing," *Cambridge University Press, Cambridge*, 1992.

[7] V. Arsigny, O. Commonwick, X. Pennec, and N. Ayache, "Statistics on diffeomorphisms in a Log-Euclidean framework," *MICCAI 2006, LNCS, Springer-Verlag, Berlin, Germany*, vol. 4190, pp. 924 – 931, 2006.

[8] J. Nocedal and S. J. Wright, "Numerical optimization," *Springer Verlag, New York*, 1999.